
Berit Gerritzen¹, Gebhard Kirchgässner²*

¹Swiss Institute for International Economics and Applied Economic Research, University of St. Gallen, St. Gallen, Switzerland
²Swiss Institute for International Economics and Applied Economic Research, Leopoldina, CESifo and CREMA, University of St. Gallen, St. Gallen, Switzerland
Email: Gebhard.Kirchgessner@unisg.ch, berit.gerritzen@gmail.com

Received 11 May 2016; accepted 25 June 2016; published 28 June 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution-NonCommercial International License (CC BY-NC).
http://creativecommons.org/licenses/by-nc/4.0/

Abstract

Given that as a whole the literature on the deterrent effect of capital punishment is inconclusive, the fact that individual authors persistently claim to have found solid evidence in one or the other direction raises two questions. Firstly, what are the causes of these different results? Do different data samples, estimation methods or time periods lead to different results or do the outcomes merely reflect prior convictions on the part of the authors? Secondly, to what extent is it possible to derive such divergent results by slightly changing the specification of the test equations without violating scientific standards? After conducting a survey of the more than forty available reviews of this literature, we present a meta-analysis of 109 deterrence studies published between 1975 and 2013. The only statistically significant explanatory variable in these studies turned out to be the profession of the author: Economists claimed significantly more often than members of legal or other social science departments to have found a significant deterrent effect. Furthermore, using a panel data set of U.S. states, we show how easy it is to derive contradictory results by employing alternative specifications. Thus, our results reinforce the claim that the empirical evidence presented to date is far too fragile to provide a basis for political decisions.

Keywords

Death Penalty, Deterrence, Econometric Evidence, Ideology

*Corresponding author.

1. Introduction

Ever since the publication of Ehrlich’s seminal paper in 1975 [1] on the deterrent effect of the death penalty, in which he claimed that each execution in the U.S. may prevent up to eight murders, there has been an intense and ongoing debate (mainly) in the United States about whether or not such a deterrent effect exists. Some of the scholarly contributions to this field have played a role in testimony given before congressional committees [3] [4]. Despite the use of modern econometric techniques, to date no consensus has been reached among scholars participating in this debate, who include not only economists but also researchers from legal and other social science departments. This paper aims to shed some light on the underlying reasons for these contradictory results, reasons that have thus far remained unclear.

We can distinguish between three generations of literature on the death penalty. The first generation comprises contributions that were published between 1975 and 1978, i.e., around the time when the moratorium on capital punishment was lifted in the United States, while the second one comprises papers written after 1982, covering the time after the moratorium ([5], p. 197f.). These earlier studies typically relied on time series and cross-sectional data, and, except for a few studies that looked at Canada or the United Kingdom, most analyses used evidence from U.S. crime statistics. In recent years, a third generation of papers has been published. The main difference between this work and the two previous generations is that more recent publications exhibit an increasing reliance on the use of panel data, mostly for U.S. states, but partly also for U.S. counties. Nevertheless, despite the vast number of analyses carried out and the similarity of the underlying data, the results are rather mixed and still far from conclusive.

The debate on the deterrence effect of the death penalty is characterized by a rather pronounced divide between academic disciplines, with (U.S.) economists much more likely to support the deterrence hypothesis than other social scientists, particularly legal academics. However, in contrast to earlier discussions, the econometric methods used by the two groups are now practically identical: With empirical rigor becoming increasingly important in all fields of social science, many non-economists are now able to use advanced statistical techniques just as well as their economist colleagues. This has in turn led to more widespread discussion of methodological problems, in particular with respect to the reliability of data and the quality of the instruments used in instrumental variables estimates.

The question of deterrence has now been a subject of economic research for more than four decades and has resulted in over 150 academic publications in this field. However, despite the large number of econometric analyses undertaken, research into the effect of the death penalty on the behavior of potential offenders remains far from conclusive. The import of disagreement among scholars in this area is increased by the fact that the subject of deterrence is of high policy relevance. Since research in this area is likely to have an influence on policy decisions, it is important whether the results suggest the existence of a deterrence effect or not. Furthermore, policymakers often lack rigorous training in econometric analysis and might therefore not be able to assess the quality of empirical studies. Based on these considerations, the National Research Council (NRC) already conducted a study in 1978 that aimed to “provide an objective assessment of the scientific validity of the technical evidence, focusing on both the existence and the magnitude of any crime-reducing effects” ([14], p. vii). The resulting report concluded that “available studies provide no useful evidence on the deterrent effect of capital punishment” (p. 9). The NRC report further suggested that more sophisticated econometric methods as well as improved data—namely more detailed and/or disaggregated data—would be required before a definitive stance could be taken on the matter (pp. 12ff). In the decades since the first NRC report, efforts have been undertaken in both these directions.

However, nearly forty years later, the academic debate on the issue of deterrence is far from being settled. A second study by the National Research Council conducted in 2012 concluded that “research to date on the effect of capital punishment on homicide is not informative about whether capital punishment decreases, increases, or has no effect on homicide rates. Therefore, the committee recommends that these studies not be used to inform deliberations requiring judgments about the effect of the death penalty on homicide. Consequently, claims that research demonstrates that capital punishment decreases or increases the homicide rate by a specified amount or

1The paper by Ehrlich was written in response to a book by Sellin in 1959 [2].
2For Canadian studies see, for example, [5] and [7], for U.K. studies see [8] and [9].
3For an overview of this literature see, for example, [10] as well as [3] and [4] and the discussion in Section 2.2 below.
4In recent years, there have been only a few studies using cross-section or time series data. For example, see [11] and [12].
5See, for example, [13] and [10].
B. Gerritzen, G. Kirchgässner

has no effect on the homicide rate should not influence policy judgments about capital punishment” ([15], p. 12).

Given this situation, one could easily gain the impression that the economics of crime and the question whether the death penalty deters murder or not amount to nothing more than a playground for ideologists. The notion that prior beliefs might have an impact on the reported results on the deterrent effect of the death penalty has already been discussed by McManus [16] in 1985. He states that divergent outcomes do not necessarily imply that authors are shirking; rather, they might be the result of selective perceptions: if contradictory results can be derived using appropriate econometric techniques, authors tend to choose those results that are in line with their a priori convictions and look for strong arguments in order to support them.

If we accept that overall the literature is inconclusive, the fact that individual authors persistently claim to have found solid evidence in one or the other direction raises two questions. Firstly, what are the causes for these different results? Do different data samples, estimation methods or time periods lead to different results or, following McManus [16], do the outcomes merely reflect prior convictions on the part of the authors? Secondly, to what extent is it possible to derive such divergent results by slightly changing the specification of the test equations without violating scientific standards? Do both sides have plausible arguments for their respective specifications? To answer the first question, we present a meta-analysis of 109 papers in order to investigate the causes for the different results (Section 3). Thus, we are the first one that not only suggest this proposition but test is using modern econometric methods. In doing so, we have attempted to include all previously published papers on the issue of deterrence that employ statistical or econometric procedures in their analysis. To answer the second question, we use a data set originally employed in 2006 by Dezhbakhsh and Shepherd [18] and subsequently by Donohue and Wolters [19] to show how easy it is to generate different results (Section 4). We thus contribute to the literature by both investigating the underlying reasons for contradictory results and by illustrating that contradictory results need not reflect dubious academic practices. Rather, the sensitivity of the results supports the claim by the National Research Council that the empirical evidence regarding the deterrence effect of the death penalty is far too fragile to provide a basis for political decisions. Prior to all this, we present a survey of the over forty reviews of the deterrence literature published to date (Section 2) in order to set forth the central issues addressed in the discussion among different (groups of) authors.

2. A Survey of Previous Reviews

Generally speaking, one can distinguish between two types of literature review: systematic reviews and meta-analyses. The former aim to provide an overview of the existing literature in a certain field, whereas the latter can have two different aims: i) to more closely identify the true value of a parameter of interest by pooling results of previous studies, and ii) to investigate what causes the different results. These two aims cannot always be reconciled with one another. In the following, we first consider the many systematic reviews of the deterrence literature and then discuss several meta-analyses that have been carried out in this field. Unless otherwise indicated, the reviews refer to studies that were conducted using U.S. data.

Many reviews of the literature on the deterrence effect of the death penalty tend to have a strong focus on methodological issues and in particular on the extent to which the use of different estimation methods can lead to different results. Given that the issue of deterrence cannot be analyzed using an experimental approach, researchers have to test their hypotheses using historical data. However, since it is not clear from the outset what a correct model for testing deterrence would look like, researchers need to formulate arguments as to why a particular type of model setup is appropriate for testing the deterrence hypothesis. Fagan [20] therefore claims that model uncertainty is intrinsic to such studies.

2.1. Reviews of First and Second Generation Studies

The first and second generation of economic research on the deterrence effect of the death penalty comprise publications that fall into the time period between 1975 and 1978 and between 1982 and 2000 respectively. These analyses mostly rely on time series data from the U.S. in order to estimate a murder supply function, although some authors also use cross-sectional data. In their comparisons of earlier studies, both Cameron ([5], pp. 197ff.) and Klein et al. ([21], pp. 146ff.) point to the sensitivity of results to the inclusion of variables such as gun ownership, other crime rates, incarceration terms and dummies for executing states. In an earlier review,

---

6For our search, besides others, we primarily used EBSCO and Google Scholar. Preliminary results of this research were presented at the Thünen Lecture at the 2012 Annual Meeting of the Verein für Socialpolitik in Göttingen, Germany. See Kirchgässner [17].
Gibbs ([22], pp. 304f.) also criticizes the fact that the inclusion of extra-legal correlates of crime (such as the unemployment rate or the proportion of non-whites in the population) is rarely based on a well-defined theoretical foundation; they employ generative and inhibitory classes of extra-legal variables, but the selection of such variables often remains obscure. (See also [23], pp. 183ff.)

Cameron [5] also points out the extent to which results depend on the choice of the functional form, i.e., linear vs. multiplicative or log-linearized specifications. The results achieved by Ehrlich [1] appear to hinge on a logarithmic transformation of the data, an argument first put forward by Bowers and Pierce ([24], p. 206) and subsequently by Passell and Taylor [25]. (See also [26], p. 7.)

A more fundamental concern regarding econometric analysis based on the Ehrlich paradigm is related to the potential for simultaneity bias. According to Klein et al. ([21], pp. 144ff.), many variables in the murder equation are mutually dependent. For example, it is hardly possible to argue that, given the probability of arrest, the probabilities of conviction and execution are exogenous variables. Yet even when a two-stage least squares approach is employed, as in the case of Ehrlich [1], this problem is not fully resolved. This is because in such models economic factors are assumed to affect criminal behavior, but not the other way around. (See also [21], p. 145, p. 151.)

Furthermore, in a review of first generation studies, Barnett ([27], pp. 364ff.) points out that most of these papers rely on the assumption that the variance of the error in estimating homicide rates is either the same in all states or inversely proportional to a state’s urban population. He re-estimates these studies and his results suggest that using weighted least squares is more appropriate when it comes to weighing state-specific data and achieving homoscedasticity. In addition, serial correlation in the stochastic part of the murder supply function needs to be taken into account by using standard errors that have been adjusted for autocorrelation. (See again [21], p. 145, p. 151.)

2.2. Reviews of Third Generation Studies

The third generation of deterrence literature uses the natural variation resulting from the fact that not all states restored capital punishment at the same time when the constitutional barrier was lifted in 1976. Most of these papers were written after the year 2000 and rely on state- or county-level data and panel estimation techniques. Besides using highly comparable data, these studies are also based on relatively similar conceptualizations of criminal behavior. However, they arrive at very different results regarding the existence of a deterrence effect.7

In a review of third generation studies, Cohen-Cole et al. ([32], p. 337) argue that the differences between the outcomes of these studies are mainly due to different model specifications and the assumptions linked to them and that a relatively small difference in the covariates included or the econometric method employed can lead to fundamentally different results. Kirchgässner ([10], pp. 461ff.) also critically reviews deterrence papers written in the last ten years and shows how simple changes in methodology and model specification can produce completely different results from the same data. Although third generation papers use more sophisticated methodology than earlier research, many problems persist, particularly their failure to adequately address issues relating to simultaneity. In contrast to earlier research, however, functional form does not appear to be a major concern in third generation literature; log-linear specifications lead to results similar to those produced by linear models.

Donohue and Wolfers [19] compare panel data studies by Katz et al. [19], Dezhbakhsh and Shepherd [18], Mocan and Gittings [28], and others. Besides providing a critical review of the evidence, they also use the data presented in these studies to re-estimate their equations. The authors find that the results produced by Dezhbakhsh and Shepherd [18] are sensitive to both the exclusion of Texas and the definition of the execution variable. In this paper, the execution variable is defined as the number of executions in a given state without controlling for population size. When scaling the execution variable per 100,000 residents, the effect becomes insignificant.8 In their review of Mocan and Gittings [28], Donohue and Wolfers use lagged values of one year instead of seven when constructing the deterrence variables, thereby rendering the coefficient insignificant ([19], p. 816ff.). The authors further show how sensitive the results produced by Dezhbakhsh and Shepherd [18] are to

7These studies include Dezhbakhsh and Shepherd [18] and Mocan and Gittings [28], who find a deterrent effect, as well as Donohue and Wolfers [19], Fagan [4] [20] and Katz et al. [29], who argue that the evidence is too sensitive with respect to model specification to draw clear conclusions. Shepherd [30] presents evidence for both the deterrence and the brutalization hypothesis, and tries to explain the conditions under which these hypotheses find support. [31] provides an extended review of this literature.

8See Donohue and J. Wolfers ([19], pp. 813ff.). See also the replication of some of their results in [33]. For a similar sensitivity test of the newer deterrence literature, see also Fagan ([20], pp. 308-311).
the instrumental variable definition. For example, Donohue and Wolters [19] find that each execution costs more than eighteen lives rather than saving eighteen lives when a minor change is made to the coding of the instrumental variable measuring partisan influence, i.e., the state level Republican vote share in the most recent presidential election.9

2.3. Heterogeneity and Model Specification

Among other things, model specification reflects what one considers to be an appropriate way of dealing with observed and unobserved heterogeneity. Keckler [34] provides an overview of both theoretical models and empirical studies in the context of heterogeneity and concludes that if there were a deterrence effect, one would expect this effect to differ between different groups of criminals. In particular, the deterrent effect of the death penalty should be more pronounced for criminals who respond less to other forms of deterrence (such as prolonged prison sentences) because of their particular cost-benefit considerations ([34], pp. 116ff.). More specifically, this would apply to gangs, serial killers, terrorists and so forth.10

Also addressing the issue of heterogeneity, Cook ([37], p. 241, p 252) and Fagan ([20], pp. 276ff.) criticize the fact that most deterrence studies do not distinguish between different types of homicides. By combining, for example, crimes of jealousy with more or less rationally planned killings, one implicitly assumes that all forms of homicide are equally deterrable.11 This is particularly questionable in light of the fact that the law explicitly distinguishes between different types of homicide and, as a result, if felony murder rules are applied, the probability of a capital punishment ruling will vary between different types of murders.

Besides focusing on heterogeneity with respect to groups of criminals and types of crimes, several review studies also focus on geographic, and in particular state-specific, heterogeneity. Donohue and Wolters ([19], pp. 826ff.) show how sensitive results are to the exclusion of very active states: if Texas and California are dropped from the data, the estimated effects in terms of lives saved or lost range from −42 to +34 and −29 to +30 respectively. Again, this can be taken as evidence of the fragility of the results.12

Furthermore, according to Kirchgässner ([10], pp. 467ff.), the failure to adequately consider heterogeneity raises concerns relating to simultaneity. Provided that a potential criminal can simultaneously choose between different criminal acts, the most appropriate way of dealing with this simultaneity would be to estimate a system of equations rather than only one equation for murder rates. However, a large part of the deterrence literature continues to rely on only one murder supply equation. Another simultaneity problem to which the literature has so far failed to respond appropriately is linked to state-specific heterogeneity. Specifically, states with lower murder rates will have less need for the use of the death penalty. On the other hand, their murder rate might simply be lower due to the fact that by performing fewer executions the state has experienced less brutalization.

2.4. Importance of Particular Assumptions

Another relevant concern of previous review articles relates to the assumptions made by economic analyses of the deterrence effect of the death penalty. Specifically, economic models of crime assume that criminals are rationally behaving individuals and that the emotionality of murder does not constitute a barrier to deterrence.13 The literature also assumes that the subjective risk perception of a potential criminal individual is equivalent to the objective risk of apprehension, conviction and execution.14 Garoupa [44] has reviewed behavioral economic alternatives to the classical economics-of-crime approach. According to him, relaxing the rationality assumption would allow for an enrichment of modelling behavior, but comes at the cost of losing tractability (p. 12). Anoth-

---

9 Instead of using six variables based on the vote share in each of the six separate presidential elections, as in the case of Dezhbakhsh and Shepherd [18], Donohue and Wolters merge this information into one partisanship variable ([19], pp. 821-825).
10 See also Donohue ([35], p. 796ff.). In this context, it is also worth noting that around the year 2000, violent death rates among street gang members in the U.S. were close to seven per cent, whereas in the same time period three percent of death row inmates were executed. In other words, for particular types of criminals, the risk of violent death appears to be higher than the risk of execution. See for this Levitt and Miles ([36], pp. 156ff.).
11 A notable exception in this regard is Shepherd [38].
12 This particular estimate was performed with the data previously used by Dezhbakhsh, Rubin and Shepherd [39]. Similar results have been achieved by Fagan [4], as well as Weisberg ([40], p. 159).
13 For critiques, see ([26], p. 5) and ([41], pp. 254ff.). On the rationality assumption in the economics of crime literature see also ([42], pp. 4ff.), as well as [43].
14 See ([20], pp. 292ff.) and ([22], pp. 281ff.). The fact that those states without the death penalty have—on average—lower homicide rates supports the existence of such reversed causality.
er critique has to do with the assumed direction of causality, with most analyses assuming that an increase in the risk of execution causes a change in the murder rate, although it might just as well be the other way around. (See [26], p. 5 and [37], pp. 255ff.)

Implicit and explicit assumptions regarding time- and state-specific heterogeneity are also mentioned as a concern in several reviews. For example, a time series approach such as that used by Ehrlich [1] assumes that the way in which the covariates influence murder rates does not change over time. However, it is difficult to explain why, for example, the impact of the share of unemployed or non-white members of the population should remain unaffected by the pronounced changes in social and civil rights legislation introduced between the 1930s and the 1960s. (See [41], pp. 248ff. and [45], pp. 294ff.)

According to Manski and Pepper [46], conclusive results on the deterrent effect of the death penalty can only be achieved if one is willing to make rather strong assumptions. The authors start with a set of weakest restrictions and find very ambiguous results. The stronger the assumptions they impose, the smaller the degree of ambiguity. However, since the gain in conclusiveness hinges on the introduction of strong assumptions and therefore the results are very sensitive, it is questionable how much knowledge one gains from this. Similar to other reviews, Manski and Pepper [45] provide evidence for diametrical results depending on the assumptions they impose.

2.5. Prior Beliefs of the Researcher

Using Bayesian statistical methods, McManus [16] shows how prior beliefs of the researcher and selective perceptions can influence results in deterrence studies. Given that the deterrence question cannot be answered in an experimental setting, the researcher has to decide which of many different specifications he is going to select for publication. Faced with contradictory results, researchers might tend to pick the specification that leads to a result close to their prior hypothesis. In other words, researchers could be tempted to try to find convincing support for the position that is the most aligned with their prior beliefs. According to McManus ([16], p. 425) this issue cannot be resolved unless one is willing to abandon the single-equation framework used in most deterrence studies. (See also [10], p. 451.)

In the case of the death penalty, results appear to be influenced not only by ideology and prior beliefs, but also by the researcher’s particular discipline. Economists, in particular, are much more likely to find evidence for a deterrence effect than other social scientists, even when using the same data sets as the researchers from disciplines such as law or sociology. (See also [13], p. 798.) Kirchgässner ([10], p. 469) explains this by pointing to the fact that “economists believe in incentives more than other social scientists do”. Similarly, Cameron ([47], p. 308) asks, “Why are economists so keen to endorse deterrence when they could reasonably suppose that it doesn’t work? Prior social conditioning could be responsible as economists grow up in a particular culture and are no more immune to its myths than anyone else.”

2.6. Data Quality

Most systematic reviews dedicate a section to data quality issues. Early deterrence research relied heavily on the use of national time series data from the FBI based on the Uniform Crime Reporting Scheme (UCR). The use of this data is problematic insofar as the UCR scheme was voluntary in its early years and many agencies did not participate. After 1960, the number of agencies that complied with the UCR and reported arrests and convictions increased dramatically. (See [26], pp. 7ff.) Homicide figures were adjusted ex post by the FBI based on current data, although authors such as Bowers and Pierce ([24], p. 190) have raised doubts regarding the validity of these estimates. Starting with Layson [48], researchers in recent decades have referred to the National Vital Statistics System (NVSS), which can be regarded as providing more reliable homicide figures. (See [49], pp. 35ff.)

Nagin [50] reviews 23 studies, most of which rely on the use of the National Prisoner Statistics (NPS) or the Uniform Crime Records (UCR). Although at first sight the evidence could be interpreted as supporting the deterrence hypothesis, he demonstrates that there are substantial flaws in the data ([49], pp. 111ff.). He shows that measurement error (intended or unintended distortions in crime data either across jurisdictions or across time) can generate an inverse association between published crime rates and any sanction variable that has published crime rates in its denominator ([50], p. 97, p. 113). For example, states where police departments record fewer crimes (than actually occur) will then tend to have lower crime rates and higher measures of sanction rates—which could create the illusion of deterrence. (See [51], but also [47], p. 308ff., [26], pp. 5ff., [23], pp. 187ff., as
well as [49], pp. 38f.) Since in many deterrence studies the denominator in the execution rate is the same as the numerator in the conviction rate, the potential bias resulting from measurement error ought to be taken very seriously. See also ([5], p. 204). Furthermore, Nagin ([50], p. 98, p. 129) points out that in jurisdictions with tougher sanctions (i.e., in states where conviction probability is higher and/or the time served is longer), the inverse association between crime and sanctions is likely to reflect an incapacitation effect rather than a deterrence effect. This is because during the time a criminal individual is incarcerated he cannot commit another crime. So naturally, states that keep a larger share of their (potentially criminal) population locked up in prisons will reduce their crime rate. In this situation, confounding incapacitation and deterrence effects will also bias results. (See also [22], pp. 293f. and [50], p. 98.)

According to Kirchgässner ([10], p. 466) data issues persist even in more recent, methodologically more sophisticated studies. In particular, the results are still extremely sensitive to the definition of the deterrence measure and the dependent variable. Furthermore, data on important states such as Florida is still far from complete, which, according to Fagan [4] and Weisberg ([40], pp. 159f.), is also likely to bias results.

In addition, several reviews point to problems linked with the use of nationally aggregated data and specifically with causal inferences regarding the deterrent effect of execution in one state on murder in another. See ([52], pp. 46f.). An alternative approach involves performing time series analyses on individual states, although it remains unclear to what extent the results can be applied to other states. (See [53], pp. 175ff., [49], pp. 39f. and [41], pp. 248f.) Using cross-sectional data on all U.S. states alleviates this problem to some extent, but, as Chan and Oxley ([26], p. 8) as well as Cook ([37], pp. 256f.) point out, it introduces difficulties related to unobserved heterogeneity. This problem is alleviated in newer studies that mainly rely on panel data and state-specific fixed effects, often combined with an instrumental variable approach such as found in Shepherd ([38], pp. 6f.). However, as Donohue and Wolfers ([19], pp. 804f.) argue, these studies remain very sensitive to changes in the model specification and/or the construction of the instrumental variables.

### 2.7. Sample Period

Cameron ([5], p. 204) presents numerous studies that show that the identification of a deterrence effect crucially depends on the exclusion of post-1962 data in the U.S. and of data from the years 1956-1968 in the U.K. See also ([49], pp. 36-38), as well as ([21], pp. 146-147). The combination of execution-free data with earlier data has been criticized in several review studies, for example by Baldus and Cole ([53]: pp. 184f.), Cameron ([47], p. 209) and Glaser ([41], pp. 243f.). Cameron ([5] argues that since there is evidence of a structural break in the data, pooling the two series would not be appropriate.

In a review of the new deterrence literature using panel data, Fagan ([20], pp. 284ff.) shows that the results are very sensitive to extensions in the observation window. According to him, this is mainly due to the fact that while the number of executions sharply decreased from 1999 to 2004, the murder rate remained relatively stable.16

### 2.8. Meta-Analyses

As mentioned above, it is not clear from the outset what a correct model for testing deterrence would look like, and researchers thus need to explain why a particular type of model setup is appropriate for testing the deterrence hypothesis. Needless to say, this entails the risk of subjective beliefs and ideology influencing the choice of data or the estimation technique. Combining the results of different studies into one meta-study therefore represents one way of reaching results that are more objective. Yang and Lester ([54]) have carried out such a meta-analysis covering 104 peer-reviewed journal articles produced following the publication of Ehrlich [1]. Only 95 studies are considered by Yang and Lester ([54], pp. 457ff.) to be based on adequate data. According to their evaluation, 60 of these studies provide evidence for the deterrence theory and 35 provide evidence against it. However, the results appear to hinge on the type of study carried out: Time series and panel data studies on average find a deterrence effect. The results from cross-sectional studies, studies of single executions and those taking into account the publicization of executions, however, were inconclusive, i.e., the average effect was not statistically different from zero.

---

15The sensitivity of Ehrlich’s results with respect to the sample period and the choice of functional form was first noted by Bowers and Pierce [24] and subsequently by Passell and Taylor [25].
16Kirchgässner [10] also discusses the sensitivity of the results with respect to the choice of the sample period.
In the above-mentioned literature review of third generation studies, Cohen-Cole et al. ([32], p. 337) argue that relatively small methodological differences can lead to fundamentally different results. One can find theoretical support for both assumptions and both statistical methods, which makes it very challenging to compare such studies, as they are both “right” in some way. In order to deal with this model uncertainty, Cohen-Cole et al. ([32], pp. 338ff.) take a model-averaging approach and provide averaging model-specific estimates using posterior probabilities. Their main conclusion reached (p. 364) is that there is a large but imprecise deterrence effect and that the strong results presented, for example, by Dezhbakhsh et al. [39] are mainly due to particular model choices.

In a meta-analysis of eighteen third-generation deterrence studies, Kirchgässner ([10], pp. 466ff.) makes a similar point. Given that the results of the included studies differ widely, it is difficult to draw firm conclusions. Overall, the mean of the reported t-statistics is negative, which provides some support for the deterrence hypothesis. However, the evidence remains inconclusive, particularly because of the sensitivity of the results to specification changes.

Dölling et al. ([56], pp. 209f.) have performed a meta-analysis of the deterrence effects of different types of punishment. In total, 391 studies are considered in this meta-analysis, of which fifty-two studies are concerned with the death penalty. The authors find a significant deterrence effect of punishment in cases of minor crime, particularly in studies relying on experimental data. However, the meta-analysis does not indicate that the death penalty deters (pp. 219-221). Again, there appears to be a pronounced influence of methodological and statistical choices on the results.

A comparable set of studies has been analyzed by Hermann [13], who conducts a meta-analysis of eighty-two papers on crime and punishment, of which fifty-two study the effect of the death penalty. Among the death penalty studies, he finds a pronounced divide between economics and other social science disciplines with respect to the results of deterrence studies. Forty-nine per cent of publications by economists in economics journals find a statistically significant deterrent effect. However, only twenty-eight percent of the publications by other social scientists, such as criminologists, sociologists, and legal academics argue that the death penalty has a statistically significant deterrent effect. He also shows that there is a correlation between the publication medium and the result of a study, namely that papers by economists who claim to prove the existence of a deterrence effect are mainly published in economics journals. However, he does not employ formal statistical tests. It therefore remains unclear whether this latter effect simply reflects the fact that economists mainly publish in economics journals or whether the editors of economics journals exercise an independent influence, for example, by preferring papers that support the existence of a deterrence effect.

Two of the available meta-analyses, those by Hermann [13] and Kirchgässner [10], thus provide some evidence that the profession of the authors might have an effect on the presented results. However, neither paper employs formal statistical tests in order to assess this hypothesis. In addition, the results achieved by Yang and Lester [54] suggest that time series analyses are more likely to produce a significant deterrence effect than studies using cross-sectional data. However, this finding has also not been statistically tested. There is thus scope for a further meta-analysis that investigates these hypotheses in more detail. Moreover, such a meta-analysis can take advantage of the fact that the available data set is larger than the ones used in these earlier studies.

3. A New Meta-Analysis

We have tried to include in this meta-analysis all papers containing original analyses published over a period starting with the paper by Ehrlich [1] and ending in 2013. Altogether, we have found 109 papers employing statistical procedures, of which 92 come to definite conclusions. In 37 papers, results in favor of a significant deterrence effect are presented that might justify the imposition of the death penalty. 55 papers conclude that if there is any evidence for a deterrent effect at all, it is so precarious that the imposition of the death penalty cannot be justified by the results of empirical research. The remaining 17 papers are inconclusive, i.e., the authors did not want to derive any conclusion or policy-recommendation from their analysis. The literature review presented above as well as the results of the previous meta-analysis by Hermann [13] suggest that a major cause of the differences in results may have to do with whether the authors of the studies are economists or not. How-

---

17 A different approach has been proposed by Leamer [55]. See also ([32], p. 339).
18 On average, the reported t-statistics are ~0.78 with a standard deviation of 5.710 (see [10], p. 466).
19 Preliminary results of this study can be found in [57]-[60].
20 The classification of the corresponding papers is given in Table A1 of Appendix.
ever, other factors might also play a role, such as the kind of data used—as pointed out, for example, by Yang and Lester [54]—and differences between methods and estimation procedures. Moreover, given that the models attempt to explain the number of homicides in a certain area and year, economic conditions might also have an effect. Finally, if not only economists participating in this debate but also editors of economics journals have stronger a priori beliefs in favor of the deterrent effect of death penalty than, for example, editors of sociological or law journals, it might be easier to publish results supporting a deterrent effect in an economics than, for example, in a law journal. If this is the case, we are dealing with reverse causality, but this should also become obvious in the estimation equation of a meta-analysis.

We start with the following equation:

\[ \text{DET} = f (\text{AECON}, \text{ANECON}, \text{ECONJ}, \text{NECONJ}, \text{TS}, \text{CS}, \text{OLS}, \text{INST}, \text{WLS}, \text{US}, \text{YEAR}, \text{UER}, \text{Y}) \] (1)

with:
- DET The study claims that there is a significant deterrence effect.
- AECON The (one) author is an economist.
- ANECON The (one) author is not an economist.
- ECONJ The paper has been published in a (scholarly) economics journal.
- NECONJ The paper has been published in a scholarly journal, but not in an economics one.
- TS Time series data are used.
- CS Cross-sectional data are used.
- OLS The equation is estimated using OLS.
- INST More advanced estimation methods (IV estimators, GMM) are employed.
- WLS The observations are weighted.
- US The data are from the United States.
- YEAR Publication year.
- UER The unemployment rate is included in the estimation equation.
- Y The growth rate of real income is included in the estimation equation.

Because the dependent variable is a binary one, we employ a probit estimator.\(^2\) Except for YEAR, all explanatory variables are also binary variables. The standard errors are clustered according to the (groups) of authors; the 92 observations lead to 51 clusters.\(^2\)

In the empirical analysis, the dependent variable DET takes on the value “1” if a paper finds a deterrence effect, and “0” if it does not. The undecided papers will be denoted as missing. We include two dummy variables for time series (TS) and cross-sectional (CS) data; for panel data both take on the value “1”. We also include two variables for the publication medium: economics journals (ECONJ) and other (NECON) scholarly journals, such as those from the fields of sociology and law. The default category is newspaper articles. We use three variables in order to capture the methods the authors have employed: OLS for ordinary least squares, WLS for weighted estimation methods, and INST for instrumental variable estimators and other advanced estimation methods such as GMM or, for example, Poisson regression. The reference category in this case includes studies employing simple comparisons. Economic development is represented by the development of real income (Y) and unemployment (U). The publication year (YEAR) is included in order to test whether there is a trend in the results; for example, in might be that the death penalty has gained or lost acceptance in the scholarly community over time. The variable for the United States might indicate whether there are differences between the United States on the one hand and the two other countries for which studies exist, i.e., the United Kingdom and Canada.

In a first attempt, we include the variable for the author being an economist as well as the one for the author not being an economist in the equation, because there are also papers written jointly by economists and non-economists. The results are given in Table 1 (Model 1).\(^2\) The only significant variable is the dummy for cross-sectional data. Thus, with pure cross-sections and panel data there is significant less evidence for a deterrence effect than using pure time series. This corresponds to the result of Yang and Lester [52]. None of the other explanatory variables turns out to be significant. This also holds for the two variables describing the professions

\[^2\]We also used a logit estimator and an ordered probit estimator including all 109 papers, where we classified the dependent variable as “+1” if the authors confirm a deterrent effect, “−1” if they deny it and “0” if they remain undecided. The differences between the results of these estimation methods are negligible.

\[^3\]If we include all 109 papers, we get 64 clusters.

\[^4\]Preliminary results of this model are presented in [17].
Table 1. Meta-analysis of the deterrence results, 92 observations.

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters (z-values)</th>
<th>$\chi^2$ (p-values)</th>
<th>Parameters (z-values)</th>
<th>$\chi^2$ (p-values)</th>
<th>Parameters (z-values)</th>
<th>$\chi^2$ (p-values)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant term</td>
<td>-53.081 (-1.30)</td>
<td>-51.437 (-1.30)</td>
<td>-59.069 (-1.51)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author is an economist</td>
<td>1.173 (1.14)</td>
<td>1.910*** (3.16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author is not an economist</td>
<td>-1.042 (-1.02)</td>
<td>0.384</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economics Journal</td>
<td>0.346 (0.47)</td>
<td>0.550</td>
<td>1.14 (0.566)</td>
<td>0.180 (0.856)</td>
<td>0.31 (0.414)</td>
<td></td>
</tr>
<tr>
<td>Non-Economics Journal</td>
<td>0.188 (8.885)</td>
<td>0.121</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time Series Data</td>
<td>-0.236 (-0.40)</td>
<td>-0.172</td>
<td>2.85 (0.001)</td>
<td>-0.830** (0.240)</td>
<td>-0.822* (0.114)</td>
<td>4.33 (0.114)</td>
</tr>
<tr>
<td>Cross-sectional data</td>
<td>-0.941* (-2.05)</td>
<td>-0.830* (-1.66)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>-0.346 (-0.68)</td>
<td>-0.400</td>
<td>-0.263 (-0.66)</td>
<td>0.324 (0.76)</td>
<td>-0.294 (-0.53)</td>
<td></td>
</tr>
<tr>
<td>Advanced estimation methods</td>
<td>-0.282 (-0.68)</td>
<td>1.24</td>
<td>-0.279 (-0.68)</td>
<td>1.17 (0.761)</td>
<td>-0.263 (-0.66)</td>
<td>1.11 (0.775)</td>
</tr>
<tr>
<td>Weighted regressions</td>
<td>0.334 (0.79)</td>
<td>0.270</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data from the U.S.</td>
<td>0.966 (1.43)</td>
<td>0.833</td>
<td>0.923 (1.40)</td>
<td>0.923 (1.40)</td>
<td>0.923 (1.40)</td>
<td></td>
</tr>
<tr>
<td>Publication year</td>
<td>0.027 (1.27)</td>
<td>0.025</td>
<td>0.923 (1.40)</td>
<td>0.923 (1.40)</td>
<td>0.923 (1.40)</td>
<td></td>
</tr>
<tr>
<td>Unemployment rate included</td>
<td>0.088 (0.22)</td>
<td>0.042</td>
<td>0.140 (0.37)</td>
<td>0.140 (0.37)</td>
<td>0.140 (0.37)</td>
<td></td>
</tr>
<tr>
<td>Growth rate of real income</td>
<td>-0.563 (-1.26)</td>
<td>1.59</td>
<td>-0.566 (-1.27)</td>
<td>1.63 (0.444)</td>
<td>-0.444 (-1.03)</td>
<td>1.09 (0.581)</td>
</tr>
</tbody>
</table>

Overall Wald test: 10.26 (0.508) 10.15 (0.517) 10.94 (0.448)

Pseudo R²: 0.373 0.362 0.359
Standard error of the regression: 0.404 0.404 0.404
Hannan-Quinn criterion: 1.304 1.304 1.291

"***", "**", "*" or "(*)&" indicate that the corresponding null hypothesis can be rejected at the 0.1, 1, 5, or 10 per cent significance level, respectively.

The z-values of the two author variables indicate that they have a highly significant impact in their respective equations. At the same time, the value of the Hannan-Quinn information criterion does not deteriorate (Model 2) or even improves (Model 3). The variables for the kind of data are marginally significant (at the 10 percent level) in Model 3; the impact of all other variables remains insignificant, irrespective of whether we consider the single variables separately, the groups of variables or all variables together.
Besides the individual coefficients, the overall predictive quality of a model is also of interest, i.e., we would want to know whether our model is able to correctly classify the papers into the categories of “deterrence” (DET = 1) and “no deterrence” (DET = 0). The results for all models are given in Table 2. If we assume that a deterrent effect is affirmed whenever the probability estimate is equal to or higher than 0.5 and that it is denied if this probability is below 0.5, then in all there models 27 out of 36 papers that affirm a deterrent effect are correctly predicted. With respect to the 56 papers in our meta-analysis that deny the presence of a deterrent effect, our prediction is correct for 46 papers in Models 1 and 2 and for 47 papers in Model 3. Thus, in all three models close to 80 per cent of all papers are correctly classified.

For all three equations, the overall Wald test shows that all explanatory variables together—except those for the profession of the authors—do not have a significant impact. Therefore, we can reduce these equations by only including the profession of the authors. This leads to the results presented in Table 3. The pseudo R²’s are now somewhat lower, the standard errors of the regression are slightly increased, but the Hannan-Quinn information criterion has considerably improved. In Models 1a and 3a, 29 papers that affirm a deterrent effect and 44 papers that deny it are correctly predicted. Model 2a correctly classifies 30 papers that support the existence of a deterrent effect and 42 papers that deny it. This leads to a slightly lower share of correct classifications of 79.4 percent and 78.3 per cent respectively. The estimated values of the parameters of interest are, in absolute

<table>
<thead>
<tr>
<th>Table 2. Predictive quality of the meta-analyses.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>1a</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>2a</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>3a</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>5a</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>6a</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>7a</td>
</tr>
</tbody>
</table>

Models 5a, 6a, and 7a refer to the restricted specification, i.e., with an economist as author as the only explanatory variable.

<table>
<thead>
<tr>
<th>Table 3. Meta-analysis of the deterrence results, 89 observations.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Constant term</td>
</tr>
<tr>
<td>Author is an economist</td>
</tr>
<tr>
<td>Author is not an economist</td>
</tr>
<tr>
<td>Pseudo R²</td>
</tr>
<tr>
<td>Standard error of the regression</td>
</tr>
<tr>
<td>Log of the pseudo-likelihood function</td>
</tr>
<tr>
<td>Hannan-Quinn criteron</td>
</tr>
</tbody>
</table>

"***", "**", *" or "*" indicate that the corresponding null hypothesis can be rejected at the 0.1, 1, 5, or 10 per cent significance level, respectively.
terms as well as with respect to their statistical significance, quite similar. Thus, this reduced model works astonishingly well: it does not provide a perfect fit, but the professional orientation of the authors seems to be the only factor relevant to the basic message of these papers.

As the results in Table 4, Model 4 show, we get a significant result for the publication medium if we exclude the profession of the authors from the test equation. If we consider the single parameters, the estimated coefficient for the economics journals is significant at the 5 per cent level. This result holds despite the fact that the two variables for the publication media are highly correlated; the correlation coefficient is $-0.794$. The Wald test for the combined effect of both variables is, with $\chi^2 = 15.51$ and 2 degrees of freedom, significant even at the 0.001 level. This effect is, however, not robust to the inclusion of the profession of the authors, which is not only factor relevant to the basic message of these papers.

Table 4. Robustness tests.

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters (z-values)</th>
<th>$\chi^2$ (p-values)</th>
<th>Parameters (z-values)</th>
<th>$\chi^2$ (p-values)</th>
<th>Parameters (z-values)</th>
<th>$\chi^2$ (p-values)</th>
<th>Parameters (z-values)</th>
<th>$\chi^2$ (p-values)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant term</td>
<td>$-82.015^*$ (-2.47)</td>
<td>-36.164 (-1.07)</td>
<td>-45.999 (-0.91)</td>
<td>-125.88^* (-2.29)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author is an economist</td>
<td>1.302 (2.11)</td>
<td>0.500 (1.01)</td>
<td>0.823 (1.19)</td>
<td>0.486 (3.30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economics Journal</td>
<td>-0.184 (-0.30)</td>
<td>0.482 (0.55)</td>
<td>0.491 (0.38)</td>
<td>0.055 (0.667)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Economics Journal</td>
<td>-0.302 (-0.59)</td>
<td>-0.195 (0.41)</td>
<td>-0.590 (-0.71)</td>
<td>-1.132 (1.34)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time series data</td>
<td>-0.483 (-1.07)</td>
<td>-0.669 (0.34)</td>
<td>-0.027 (-0.04)</td>
<td>0.82 (0.66)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross-sectional data</td>
<td>-0.369 (-0.68)</td>
<td>-0.090 (0.49)</td>
<td>0.226 (0.85)</td>
<td>0.595 (0.85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>-0.213 (-0.58)</td>
<td>1.10 (0.777)</td>
<td>0.41 (0.939)</td>
<td>0.84 (0.839)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced estimation methods</td>
<td>0.253 (0.65)</td>
<td>0.222 (0.54)</td>
<td>-0.343 (-0.79)</td>
<td>-0.246 (-0.41)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted regressions</td>
<td>0.415 (0.59)</td>
<td>0.488 (0.94)</td>
<td>-1.489^* (-2.30)</td>
<td>-0.105 (-0.12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data from the U.S.</td>
<td>0.041^* (2.44)</td>
<td>0.017 (1.02)</td>
<td>0.024 (0.91)</td>
<td>0.063^* (2.27)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Publication year</td>
<td>0.188 (0.54)</td>
<td>-0.184 (-0.53)</td>
<td>-0.452 (-1.11)</td>
<td>-0.667 (-1.39)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployment rate included</td>
<td>-0.152 (-0.36)</td>
<td>0.125 (0.042)</td>
<td>-0.575 (0.124)</td>
<td>-0.766 (0.146)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Growth rate of real income included</td>
<td>-0.38 (0.106)</td>
<td>0.138 (0.75)</td>
<td>1.537 (1.33)</td>
<td>1.336 (1.33)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall Wald test</td>
<td>29.09^* (0.002)</td>
<td>6.99 (0.80)</td>
<td>11.53 (0.40)</td>
<td>12.69 (0.31)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^*$, $^{**}$, $^{***}$ or $^{****}$ indicate that the corresponding null hypothesis can be rejected at the 0.1, 1, 5, or 10 per cent significance level, respectively.
The distinction between papers that deny a deterrence effect and those that do not draw a conclusion might be somewhat artificial because authors who argue against this effect often do so because the evidence is too fragile to provide a basis for any serious conclusions. Therefore, in an additional equation we re-specify our dependent variable, asking only whether the authors claim that a significant deterrence effect exists or not. Thus, we get 36 papers claiming and 73 papers denying a significant deterrence effect. Results from estimating the full model but excluding the variable for authors who are not economists are given in Table 4, Model 5. We arrive at nearly the same results as in Model 2, which has the same specification but uses a restricted data set: only the variable for the profession of the author is significant. If we re-estimate the model excluding all insignificant variables, we arrive at the result that 30 of the 37 papers that claim a deterrence effect but only 46 of the 72 papers that do not are correctly predicted.\textsuperscript{24} Thus, there are now only 69.7 per cent correctly classified.

Our sample differs somewhat from the earlier sample used by Yang and Lester \cite{54}. On the one hand, we have introduced new studies that were not yet available when these authors performed their analysis. On the other hand, we have excluded some of the studies that they have taken into account because these studies, mainly older ones, did not employ econometric methods. With respect to the 75 studies included in both samples, there are also six papers where we differ with respect to the classification. This holds, for example, for the two papers by Shepherd \cite{30} and Zimmermann \cite{61}. Both authors believe in the existence of a deterrence effect but not under all circumstances. Shepherd \cite{30} finds a deterrence effect if there are “enough” executions in a state: she finds a threshold of approximately 9 executions during her sample period from 1977 to 1996. According to Zimmermann \cite{61}, the execution method matters: only electrocution is effective in this respect. Because both authors in principle believe in the existence of a deterrence effect, we have classified them accordingly, whereas Yang and Lester \cite{54} classified these papers as “undecided”.

To check whether different classifications have an impact on the results, we re-estimated our model including only those 75 papers in our sample that have also been considered by Yang and Lester \cite{54}. The results that we find when employing their classification are presented in Table 4, Model 6. The author variable for economists is still significant but only at the 5 per cent level. The dummy variable indicating that the data are from the United States is now also significant at the 5 per cent level. Taking all explanatory variables together, except for the economist as author dummy, we again fail to prove the null hypothesis that the variables jointly do not have an impact at any conventional significance level. This reinforces our result that the professions of the authors are the only variables with a significant impact on the results of the studies. If we therefore estimate the restricted model with only the profession of the author as an explanatory variable, 23 out of the 30 papers, which claim a deterrence effect and 29 of the 45 papers that do not are correctly predicted.\textsuperscript{25} Thus, 69.3 per cent of all papers are correctly classified.

If we use our own classification (Model 7), the t-statistic for an economist as author is 3.30, i.e., the estimated coefficient is again significant at the 1 per cent level. The Wald test for the combined hypotheses leads to $\chi^2 = 12.69$. With 11 degrees of freedom, this is again not significant at any conventional level: the p-value is 0.314. Now 24 papers that claim and 32 papers that do not claim a significant deterrence effect are correctly classified, which implies a correct classification of 83.6 per cent. Thus, the difference in the classification scheme has some effect, but it does not change the overall result: the only robust and significant effect is related to the profession of the author(s).

Table 2 summarizes the predictive qualities of the estimated models. Except for Models 5 and 6, in all other specifications more than 77 per cent of all cases are correctly classified. Given the fact that in nearly half of the models we only use one explanatory variable and that we use two different classifications, this is a rather high percentage rate.\textsuperscript{26}

As nearly all papers are written by American researchers, these results suggest that the majority of (American) economists who are active in this field believe that the incentives provided by the threat of the death penalty are sufficiently severe to deter a significant number of potential murderers from committing homicide. The majority of other scholars, in particular members of law and other social science faculties, appear to believe that most homicides are committed for reasons which can hardly be influenced by the threat of severe punishment. Thus, even if they exist, the incentives resulting from the threat of capital punishment are too small to justify the use of

\textsuperscript{24}This corresponds to Model 5a in Table 2.
\textsuperscript{25}This corresponds to Model 6a in Table 2.
\textsuperscript{26}Donohue and Wolfers \cite{19} are a law and an economics professor and deny a deterrent effect. Thus, the predictive quality of the estimated model is slightly higher if we employ the variable for a non-economics author instead of the one for an economics author.
the death penalty in light of its serious ethical problems. This basic difference in terms of convictions may be the reason authors come to such different conclusions even when using the same data and method. As is shown in the next section, minor changes in the estimation equations make this possible without violating the rules of serious academic work.

4. Some Re-Estimations

Deterrence research tends to produce highly contradictory results even when it is based on the same data sets and comparable methodologies. The meta-analysis carried out in the previous chapter has shown that in a comparison of 109 previously published papers on the deterrence effect of the death penalty, the only relevant predictor of different outcomes appears to be the affiliation of the authors.

If we accept the result from the previous section that economists are significantly more likely to find a deterrence effect than other social scientists, the second question stated above becomes relevant: is it possible to reach such different conclusions without violating scientific standards? As mentioned above, McManus [16] explains the divergent results as due to selective perceptions on the part of the authors based on their a priori beliefs. However, this presupposes that contradictory results can be produced by minor modifications to the estimation equations. In order to show that this is possible we employ a panel data set of U.S. states already used by Dezhbakhsh and Shepherd [18] and Donohue and Wolfers [19].

The sensitivity of death penalty data has been shown before, for example, by Katz, Levitt and Shustorovich [29], but also by Donohue and Wolfers [19]. Their approach is particular insofar as they provide exact re-estimations of original results from previous papers and subsequently show how minor modifications can alter major conclusions. We want to add to their results by subjecting an estimation strategy previously employed by Dezhbakhsh and Shepherd [18] to further scrutiny. In the first column of Table 5, we provide their original estimation results ([18], Table 8, p. 525), followed by initial modifications by Donohue and Wolfers ([19]: p. 806, p. 814, Table 5) in columns 2 and 3. The following four columns show our own manipulations of the data and estimation strategy. All estimated equations displayed in Table 5 employ weights that represent population size together with time and state-specific fixed effects. Standard errors have been clustered at the state level.

Dezhbakhsh and Shepherd [19] used a panel data set of 50 U.S. states during the period from 1960 to 2000. Their coefficient estimate for the number of executions suggests that additional executions have large and highly significant deterrence effects on the murder rate in the same state and year. As can be seen in column 1 in Table 5, the estimate of −0.145 is significant at all conventional significance levels. The execution variable used is defined as the number of executions in a given state in a given year over the time period from 1960 until 2000. The authors further use controls for per capita real income, the unemployment rate, police employment, the share of minorities in the population, and the percentage of the population that is between 15 and 19 years old and between 20 and 24 years old respectively. In addition, they employ time and state-specific control variables. In contrast to most other papers, they use decade-specific rather than year dummy variables in order to capture long-term trends in crime (p. 524).

We further extend this analysis by using the modified execution variable (i.e., the number of executions in a given state in a given year per 100,000 residents) over a longer time horizon; we extend the survey period to cover all periods from 1934 to 2000. Our estimate of the coefficient on the execution variable (column 4 in Table 5) is highly comparable to that attained by Donohue and Wolfers [19] (column 3 in Table 5), both in terms of economic as well as statistical significance. Moreover, the adjusted R² values of 0.836 and 0.835 are very similar. Thus, in the present specification, using a longer sample does not appear to affect the results; in both estimations the effect of executions on murder rates is statistically insignificant.

The estimation equation in column 5 of Table 5 also uses data from 1934 until 2000, but relies on the

---

27 We use the data as reconstructed by Donohue and Wolfers ([19], see p. 805). These data have been made available online together with the exact specifications of the estimated models. Donohue and Wolfers [19] were unable to reproduce precisely the same results as Dezhbakhsh and Shepherd [18], but the differences are negligible. (See also Footnote 43 in Donohue and Wolfers, [19], p. 805.)

28 Since both the original papers were published in the same year (2006), it should be noted in this context that the critique by Donohue and Wolfers [19] of the analysis by Dezhbakhsh and Shepherd [18] refers to the working version of their paper, which was published in 2004. However, there are no differences with respect to the results that we discuss here.

29 Donohue and Wolfers [19], Footnote 46) list a number of previous papers that use year-fixed effects rather than decade-fixed effects. Interestingly enough, the list also includes several publications by Shepherd [30] [38].

30 For this extended sample period, we use a dataset that Donohue and Wolfers [19] employed in order to re-estimate the results achieved by Katz, Levitt and Shustorovich [29].
B. Gerritzen, G. Kirchgässner

Table 5. Some re-estimations.

<table>
<thead>
<tr>
<th>Model</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant term</td>
<td>Not reported</td>
<td>−5.891</td>
<td>−4.979</td>
<td>−5.031</td>
<td>−6.006</td>
<td>−18.158***</td>
<td>1.236</td>
</tr>
<tr>
<td></td>
<td>(−1.34)</td>
<td>(−1.08)</td>
<td>(−1.07)</td>
<td>(−1.35)</td>
<td>(−3.22)</td>
<td>(0.34)</td>
<td></td>
</tr>
<tr>
<td>Executions</td>
<td>−0.145***</td>
<td>−0.138***</td>
<td>−0.140***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(−10.52)</td>
<td>(−10.71)</td>
<td>(−10.96)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Executions per 100,000</td>
<td>−8.356</td>
<td>−8.393</td>
<td></td>
<td></td>
<td></td>
<td>5.440***</td>
<td>7.395*</td>
</tr>
<tr>
<td>population</td>
<td>(−1.43)</td>
<td>(−1.40)</td>
<td></td>
<td></td>
<td></td>
<td>(1.69)</td>
<td>(2.11)</td>
</tr>
<tr>
<td>Per capita real income</td>
<td>−0.001***</td>
<td>0.000</td>
<td>−0.003</td>
<td>−0.002</td>
<td>0.001</td>
<td>0.036**</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>(−5.16)</td>
<td>(0.01)</td>
<td>(−0.22)</td>
<td>(−0.14)</td>
<td>(0.10)</td>
<td>(1.91)</td>
<td>(0.32)</td>
</tr>
<tr>
<td>Unemployment</td>
<td>−0.147***</td>
<td>−0.194***</td>
<td>−0.214***</td>
<td>−0.185**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(−2.44)</td>
<td>(−3.06)</td>
<td>(−3.36)</td>
<td>(−2.92)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Police Employment</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.63)</td>
<td>(−0.31)</td>
<td>(−0.76)</td>
<td>(−2.92)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per cent minority</td>
<td>0.167*</td>
<td>8.891†</td>
<td>9.729***</td>
<td>9.022***</td>
<td>8.118</td>
<td>17.911***</td>
<td>30.921***</td>
</tr>
<tr>
<td></td>
<td>(6.04)</td>
<td>(1.76)</td>
<td>(1.95)</td>
<td>(1.79)</td>
<td>(1.58)</td>
<td>(3.97)</td>
<td>(6.39)</td>
</tr>
<tr>
<td>Per cent 15 to 19 years</td>
<td>−1.067***</td>
<td>90.088***</td>
<td>74.598</td>
<td>77.082</td>
<td>93.655***</td>
<td>224.502***</td>
<td></td>
</tr>
<tr>
<td>old</td>
<td>(−8.38)</td>
<td>(1.78)</td>
<td>(1.44)</td>
<td>(1.44)</td>
<td>(1.81)</td>
<td>(4.94)</td>
<td></td>
</tr>
<tr>
<td>Per cent 20 to 24 years</td>
<td>2.215***</td>
<td>66.681‡</td>
<td>74.233*</td>
<td>73.075*</td>
<td>64.490†</td>
<td></td>
<td></td>
</tr>
<tr>
<td>old</td>
<td>(9.51)</td>
<td>(2.23)</td>
<td>(2.38)</td>
<td>(2.13)</td>
<td>(1.96)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prisoner death rate</td>
<td>0.046</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.73)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The numbers in parentheses are the t-statistics of the estimated parameters. ****, ***, * or (*)& indicate that the corresponding null hypothesis can be rejected at the 0.1, 1, 5, or 10 per cent significance level, respectively.

The execution variable as defined by Dezhbakhsh and Shepherd [18], i.e., the total number of executions in a given state without controlling for population size. The resulting coefficient estimate is negative, highly statistically significant, and therefore comparable to the original result that Dezhbakhsh and Shepherd [18] achieved using a shorter sample (column 1). The use of an extended sample period can be considered as a robustness test, i.e., if a relationship between two variables is only significant for a given time period, it would seem unwise to base policy recommendations on it. As the results in columns 3 and 4 as well as in columns 1, 2, and 5 of Table 5 are highly similar, this does not seem to be a problem in the case of these data. The sensitivity of the results with respect to the definition of the execution variable appears to be more worrisome. In column 6 of Table 5, we make yet another modification in which we use the number of executions per 100,000 inhabitants together with the longer sample from 1934 to 2000 but exclude the unemployment rate and the police variable. The estimated coefficient of the execution variable is now positive and statistically significant at the ten per cent level. In other words, increasing the number of executions in a given state would be associated with an increase in the murder rate of that state. This result is contrary to what Dezhbakhsh and Shepherd [18] find and yet the differences with respect to the estimation equation in column 1 and column 6 of Table 5 are neither excessive nor based on dubious scholarly practices. Relating the number of executions to population size is arguably more appropriate than using the absolute number of executions. The same holds for the use of year-specific rather than decade-specific fixed effects. (See also [19], p. 805f.) Omitting the police variable that has been insignificant across
all previously employed specifications also appears to be reasonable. The omission of the unemployment rate is, however, somewhat more difficult to justify, particularly since we find in all other estimations presented in Table 5 a statistically significant relationship between the unemployment and the murder rate. However, one could argue that per capita real income captures, to a certain extent, changes in the unemployment rate. This can also be seen in column 6 of Table 5, where the estimated coefficient on per capita real income becomes significant at the 10 per cent level once we omit the unemployment variable. Including controls for income but not for unemployment in the estimation equation is a strategy that has been employed by several authors, who notably do not agree with each other regarding the existence of a deterrence effect. (For example, see [1] [62]-[65].)

5. Concluding Remarks

Despite nearly fifty years of econometric research, there is an ongoing and lively discussion about whether or not the death penalty has a deterrent effect. It seems that many economists believe in the deterrent effect but most other scholars participating in this debate do not. Supporting this impression, our meta-analysis shows that the major and only significant driver of whether authors argue that the death penalty deters potential murderers is his/her profession. However, these divergent results do not necessarily imply that scientific standards have been violated. There are, of course, also recent studies that fall short of today’s scientific (econometric) standards, but our re-estimates using data originally employed by Dezhbakhsh and Shepherd [19] and by Katz, Levitt and Shustorovich [29] show that the use of relatively few, justifiable modifications can lead to contradictory results. This clearly shows the sensitivity of the findings and therefore supports the claim of the National Research Council in 2012 that such results ought not to be allowed to influence policy decisions regarding the appropriateness of the death penalty.31

Overall, the literature on the deterrent effect of capital punishment is inconclusive; there are serious methodological problems that are unlikely to be solved in the (near) future. As Chalfin, Haviland and Raphael in an extensive review of the more recent literature conclude: “we do not see the additional methodological tools that are likely to overcome the multiple challenges that face researchers in this domain, including the weak informativeness of the data, the lack of theory on the mechanisms involved, and the likely presence of unobserved confounders” ([31], p. 5). However, the fact that individual authors persistently claim to have found solid evidence in one or the other direction raises two questions. Firstly, what are the causes for these different findings? Do different data samples, estimation methods and time periods lead to different results, or do the outcomes merely reflect prior convictions on the part of authors? Secondly, to what extent is it possible to arrive at such divergent results by slightly changing the specification of the test equations without violating scientific standards? After a survey of the over forty reviews of this literature available to date, we performed a meta-analysis of 109 studies published between 1975 and 2013. The profession of the author is the only significant explanatory variable: Economists claim significantly more often to have found a significant deterrence effect than legal scholars and other social scientists. On the other hand, using a panel data set of U. S. states, we show how easy it is to arrive at contradictory results by employing alternative specifications. Thus, our results support the claim that the empirical evidence presented to date is by far too fragile to provide a foundation for policy decisions.

Our results also reinforce the suggestion made by McManus [16] that selective perceptions might be the cause of divergent findings. If different results can be obtained based on reasonable assumptions, researchers will consider those outcomes as reliable that correspond to their own pre-conceptions. In this respect, economists are no different from other scholars. In addition, because economists tend to believe more in incentives than other (social) scientists do, these differences in the results obtained are not really surprising.

If, following the conclusion of the National Research Council [15], the deterrence argument can no longer be used to defend the use of the death penalty; it becomes difficult to defend it at all. In this situation, moral arguments play a major role. There is, however, also a debate about the moral status of this penalty, as shown, for example, by the discussion between Sunstein and Vermeule [67] [68] and Streiker [69]. However, the former, who try to defend the death penalty for moral reasons, rely on its deterrent effect. If our results are taken seriously, this position is hard to justify. It is, of course, also possible to believe in the deterrent effect while acknowledging that the empirical evidence for it is weak. (See, for example, [70].) But then one should at least also acknowledge that there is no scientific basis for this belief.

31See also the conclusion of the recent survey by Nagin ([66], p. 102): “Studies of the deterrent effect provide no useful information on the topic.”
Acknowledgements

We are grateful to Joe O’Donnell for editing the paper in English.

References


Effects of Capital Punishment.


## Appendix

### Table A1. Classification of authors and papers.

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Affiliation(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I) Deterrent effect</strong></td>
<td></td>
</tr>
<tr>
<td>Brumm and Cloninger [71]</td>
<td>Economics</td>
</tr>
<tr>
<td>Chressanthis [72]</td>
<td>Economics</td>
</tr>
<tr>
<td>Cloninger [73] [74]</td>
<td>Economics</td>
</tr>
<tr>
<td>Cloninger and Marchesini [75] [76]</td>
<td>Economics</td>
</tr>
<tr>
<td>Dezhbakhsh. and Rubin [77]</td>
<td>Economics</td>
</tr>
<tr>
<td>Dezhbakhsh and Shepherd [18]</td>
<td>Economics</td>
</tr>
<tr>
<td>Ehrlich [1] [78] [79]</td>
<td>Economics</td>
</tr>
<tr>
<td>Ehrlich, and Liu [80]</td>
<td>Economics</td>
</tr>
<tr>
<td>Fajnzylber, Lederman and Loayza [81]</td>
<td>Economics</td>
</tr>
<tr>
<td>Land, Teske and Zeng [82] [83]</td>
<td>Sociology, Criminal Justice</td>
</tr>
<tr>
<td>Layson [7] [48] [84]</td>
<td>Economics</td>
</tr>
<tr>
<td>Liu [63]</td>
<td>Economics</td>
</tr>
<tr>
<td>Merriman [85]</td>
<td>Economics</td>
</tr>
<tr>
<td>Moean Gittings [28] [86]</td>
<td>Economics</td>
</tr>
<tr>
<td>Phillips and Hensley [87]</td>
<td>Sociology, Psychology</td>
</tr>
<tr>
<td>Shepherd [30] [38]</td>
<td>Economics</td>
</tr>
<tr>
<td>Stack [88]-[91]</td>
<td>Sociology</td>
</tr>
<tr>
<td>Wolp in [9]</td>
<td>Economics</td>
</tr>
<tr>
<td>Yang and Lester [92]</td>
<td>Economics, Psychology</td>
</tr>
<tr>
<td>Yunker [93] [94]</td>
<td>Economics</td>
</tr>
<tr>
<td>Zimmerman [61] [95] [96]</td>
<td>Economics</td>
</tr>
<tr>
<td><strong>II) No deterrent effect</strong></td>
<td></td>
</tr>
<tr>
<td>Albert [97]</td>
<td>Law</td>
</tr>
<tr>
<td>Archer Gartner and Marc [98]</td>
<td>Sociology</td>
</tr>
<tr>
<td>Avio [6] [51]</td>
<td>Economics</td>
</tr>
<tr>
<td>Bailey [99]-[112]</td>
<td>Sociology</td>
</tr>
<tr>
<td>Bailey and Lott [113]</td>
<td>Sociology</td>
</tr>
<tr>
<td>Baileyand Petersen [114] [115]</td>
<td>Sociology</td>
</tr>
<tr>
<td>Bechdolt [116]</td>
<td>Economics</td>
</tr>
<tr>
<td>Blackand Orsagh [117]</td>
<td>Economics</td>
</tr>
<tr>
<td>Bowers and Pierce [24] [118]</td>
<td>Sociology</td>
</tr>
<tr>
<td>Boyes and McPheters [119]</td>
<td>Economics</td>
</tr>
<tr>
<td>Cheatwood [120]</td>
<td>Sociology</td>
</tr>
<tr>
<td>Decker and Kohfeld [121]-[123]</td>
<td>Sociology, Criminal Justice</td>
</tr>
<tr>
<td>Dills, Miron and Summers [65]</td>
<td>Economics</td>
</tr>
<tr>
<td>Donohue and Wolfers [19]</td>
<td>Law, Economics</td>
</tr>
</tbody>
</table>
Continued

Fagen [20] Law, Public Health
Fagan, Zimring and Geller [64] Law, Public Health
Forst [62] [124] Law
Fox [125] Law
Grogger [126] Economics
Hjalmarsso [12] Economics
Katz, Levitt and Shustorovich [29] Economics
King [127] Law
Kleck [128] Sociology
Knorr [129] Law
Kovandzic, Vieraitis and Boots [130] Criminology, Criminal Justice, Psychology
Lempert [131] Law
Marvell and Moody [132] Sociology, Economics, Econometrics
Neumayer [133] Economics
Parker and Smith [134] Sociology, Psychology
Peterson and Bailey [135] Sociology
Rahav [136] Sociology
Sorensen et al. [137] Mathematics, Computer Science, Political Science, Sociology
Stack [138] Sociology
Stringham and Levendis [139] Economics
Thomson [140] [141] Sociology, Criminology

II) Undecided

Cantor and Cohen [142] Sociology
Cochran and Chamlin [143] Sociology, Criminology
Cover and Thistle [144] Economics
Durlauf, Fu and Navarro [145] [146] Economics
Ekelund et al. [147] Economics
Klein, Forst and Filatov [21] Economics
Leamer [55] Economics
Manski and Pepper [46] Economics
McFarland [148] Social Psychology
McKee and Sesnowitz [149] Economics
Passel [150] Economics
Passel and Taylor [25] Economics
Phillips [151] Sociology
Sesnowitz and McKee [152] Economics
Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, Linkedin, Twitter, etc
A wide selection of journals (inclusive of 9 subjects, more than 200 journals)
Providing a 24-hour high-quality service
User-friendly online submission system
Fair and swift peer-review system
Efficient typesetting and proofreading procedure
Display of the result of downloads and visits, as well as the number of cited articles
Maximum dissemination of your research work

Submit your manuscript at: http://papersubmission.scirp.org/